

# Thesis Notes

PAVEL STRAŽÁK

September 1, 2009

## Motto

Frasier: “How was your hunting trip?”  
Martin: “Came home empty handed.”  
Frasier: “Oh dear; I guess that means for the next several weeks we’ll hear your grouse about the grouse and carp about the carp.”  
Niles: “You’ve been working on that, haven’t you?”  
Frasier: “Well, there was traffic.”

*Frasier, Season 9, Part 3*

## Idioms

Even “non-compositional” idioms are actually (originally) metaphorical or metonymical. Even though sometimes it is hard to see that. At other times a speaker may forget that rather straightforward metaphoric aspect:

Barack Obama accused his Republican rivals of stirring a controversy over a comment he made about putting “lipstick on a pig.” (*NY Times*, 11. September 2008)

## 1 PDT 2.0

In PDT there are several functors that refer to multiword expressions (MWEs) in one way or another. There are also some technical lemmas (??? nebo jen 1?) like #Forn that identify roots of subtrees representing MWE’s.

There are currently two graphical search engines for PDT: Netgraph ? and TrEd ?. Both have their respective benefits, but since TrEd is considerably faster due to its use of an SQL database backend ?, we have used TrEd for all the examples in this work. We also give the search queries using the PML Tree Query language ? where appropriate.

## 1.1 Foreign Phrase: t-node [t\_lemma = #Forn]

Foreign Phrase seems to be overused and its overuse seems a bit arbitrary.

- jmena firem jsou nekdy forn, nekdy ne. (dohledat)

### 1.1.1 Foreign phrases with just one t-node

There are 34 occurrences of this construction in the PDT 2.0. Counting them is as easy as writing a query in Figure 1 and extending it with this filter: »count(\$n).

In case of the bibliographic reference in Figure 2 there is coordination of three foreign phrases corresponding to the parts of a bibliographic reference annotated, but the reason for this is not very clear. After all, the point of annotating foreign phrases as simple lists with a #Forn node as a head was to make no assumptions about these pieces of a text ??.

- je to kvuli te interpunkci???

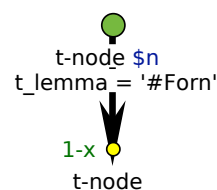
Names of companies seem to be distinguished more by the country of origin than by any linguistic reasons, as demonstrated in Figure 3. As far as linguistic criteria are concerned, Chemapol and Inekon are as foreign as Agip or Total. However the first two are, or at least were<sup>1</sup>, Czech companies, while those in the latter group have a foreign origin.

## 1.2 CPHR

There are 76 occurrences of CPHR nodes, whose head verb is not its parent, but only effective parent, in 40 sentences. See Figure 4 for the query and Figures 5 and 6 for examples.

Figure 6 shows on the other hand a coordination of two V-N idioms with the same verbal part.

Tree Query  
q-09-01-25\_000318



t-node \$n:= [ t\_lemma  
= '#Forn', 1-x t-node [ ]  
];

Figure 1: PML-TQ search query for single-node foreign phrases

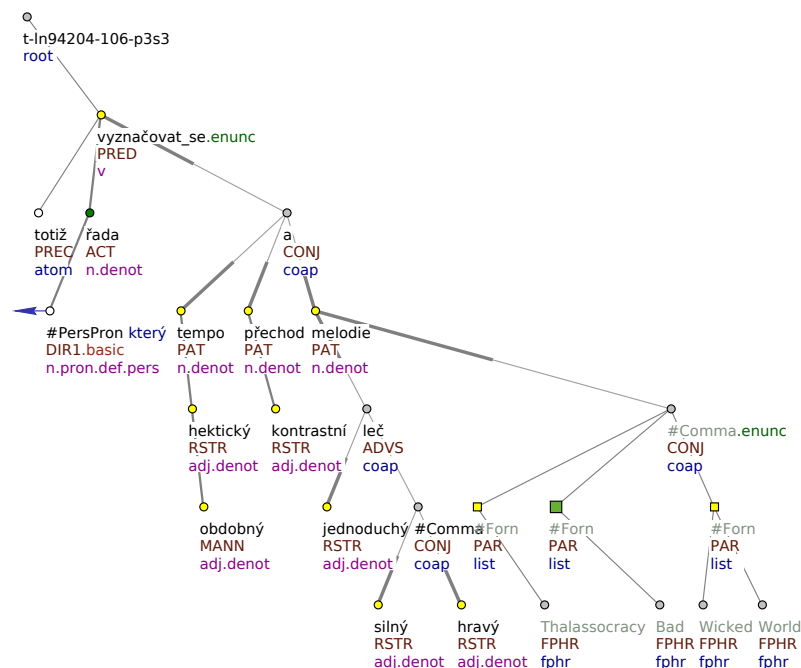
Tree Query  
q-09-01-24\_233358



t-node \$n:= [ functor = 'CPHR' ];  
t-node \$a:= [ !child \$n, echild \$n ];

Figure 4: PML-TQ search query for CPHR nodes, whose effective parent is not its parent

<sup>1</sup>At the time of writing this thesis Chemapol is owned by another international company, which only emphasises vagueness of this distinction



File: In94204\_106.t.gz, tree 7 of 17

Řada z nich se totiž vyznačuje obdobně hektickým tempem, kontrastními přechody a jednoduchými, leč silnými, hravými melodiemi (Thalassocracy, Bad, Wicked World).

Figure 2: A bibliographic reference analysed as a coordination of three foreign phrases

## 2 Notes

“*Prokopnout pětku*” – prokopnout [trestný kop after scoring] pětku – ellipsis crossing into pragmatics(?), because it is *not* an ellipsis of some exact linguistic construction, but rather an ellipsis of a situation which everyone in the discourse understand.

“Návrh / novela Zákona na ochranu osobních údajů”

((návrh ∨ novela)|zákon ∨ předpis)

– lexikální funkce?? (?) ?

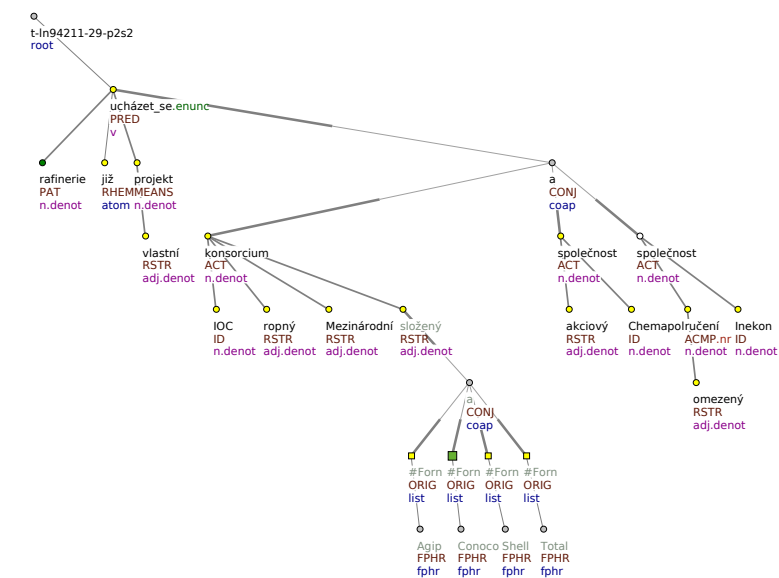
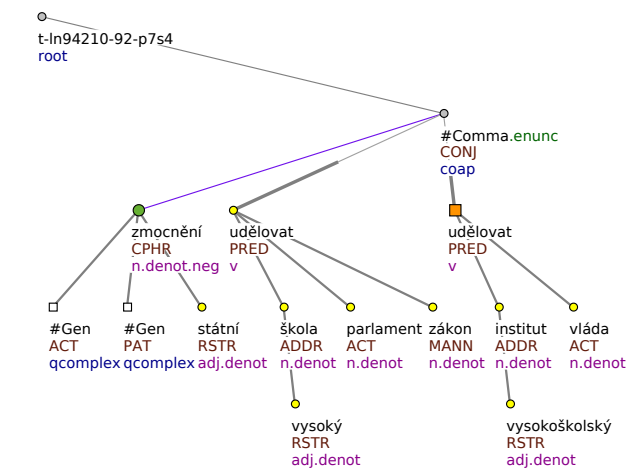


Figure 3: Annotation of Czech and foreign company names



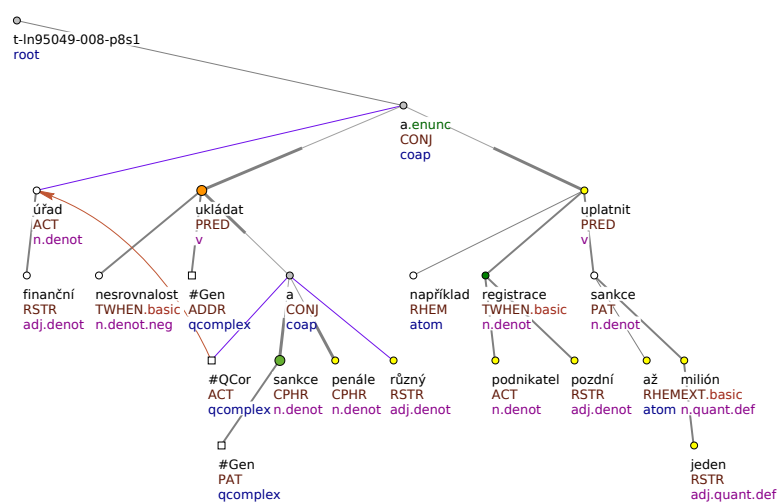
File: In94210\_92.t.gz, tree 23 of 48

Státní zmocnění by uděloval vysokým školám parlament formou zákona,  
vysokoškolským institutům vláda.

Figure 5: Coordination of verbonominal idioms, where the verbal parts are further  
???rozvite

## 2.1 Missing t-nodes in coords

“*První* a druhá světová válka” apod. – the word “*První*” is an ellipsis of “*První světová válka*”. The ellided t-nodes that should had been newly established were not, however, which left us two options, either annotate a single-word (and single-node) expression as an instance of a multi-word lexeme, or annotate the words “*světová válka*” that occure in the text as being a part of both expressions. We decided for the first, because the agreement on ellipses is fairly high in this type of coordinations.



File: In95049\_008.t.gz, tree 19 of 19

Finanční úřad může při nesrovnalostech ukládat různé sankce a penále a například při pozdní registraci podnikatele může uplatnit sankce až do jednoho miliónu.

Figure 6: Coordination of two idioms starting with the same verb